Effect of data encoding on the expressive power of variational quantum-machine-learning models

Maria Schuld,[1] Ryan Sweke,[2] and Johannes Jakob Meyer [2]

[1]*Xanadu, Toronto, Ontario, Canada M5G 2C8*
[2]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*

Quantum computers can be used for supervised learning by treating parametrized quantum circuits as models that map data inputs to predictions. While a lot of work has been done to investigate the practical implications of this approach, many important theoretical properties of these models remain unknown. Here, we investigate how the strategy with which data are encoded into the model influences the expressive power of parametrized quantum circuits as function approximators. We show that one can naturally write a quantum model as a partial Fourier series in the data, where the accessible frequencies are determined by the nature of the data-encoding gates in the circuit. By repeating simple data-encoding gates multiple times, quantum models can access increasingly rich frequency spectra. We show that there exist quantum models which can realize all possible sets of Fourier coefficients, and therefore, if the accessible frequency spectrum is asymptotically rich enough, such models are universal function approximators.

I. INTRODUCTION

A popular approach to quantum machine learning uses trainable quantum circuits as machine learning models similar to neural networks. Quantum gates—the building blocks of quantum circuits—are used to encode data inputs $x = (x_1, \ldots, x_N)$ as well as trainable weights $\theta = (\theta_1, \ldots, \theta_M)$. The circuit is measured multiple times to estimate the expectation of some observable, and the result is interpreted as a prediction. The overall computation implements a "quantum model function" $f_\theta(x)$, a machine learning model that is based on quantum computing. This approach is known by different names such as variational circuits [1,2], quantum circuit learning [3], quantum neural networks [4,5], or parametrized quantum circuits [6].

A lot of work has been done to understand the practical details of this approach, leading to useful training strategies [3,7,8] and ways to emulate and extend classical machine learning methods [2,9–12]. A growing body of literature, motivated by the dilemma of investigating the performance of quantum machine learning when only small-scale experiments are physically possible, tries to understand the potential power of variational circuits from a theoretical perspective [5,13–15]. Still, only little is known about the actual *function classes* that quantum circuits give rise to. Can quantum models express any function in the input x, or are they limited to a specific class of functions? Can this class of "learnable functions" be characterized in a meaningful way, and can the characterization be used to guide design choices and potential applications for these quantum models?

In this paper we investigate these questions in a framework focused on the role of data encoding. We consider standard models from the literature that consist of multiple "circuit layers," each made up of a *data-encoding (circuit) block* and a trainable (circuit) block, and assume that input features $x \in \mathbb{R}$ are encoded by gates of the form e^{ixH}, where H is an arbitrary Hamiltonian. Our main tool is the natural representation of such quantum models as a Fourier-type sum

$$f_\theta(x) = \sum_{\omega \in \Omega} c_\omega(\theta) e^{i\omega x}, \tag{1}$$

where ωx is the inner product. We show that the *frequency spectrum* $\Omega \subset \mathbb{R}^N$ is solely determined by the eigenvalues of the data-encoding Hamiltonians while the design of the entire circuit controls the coefficients c_ω that a quantum model can realize (see Fig. 1). The representation of quantum models as Fourier-type sums characterizes the function families that a given class of quantum models can learn via two interrelated properties. The first property is the frequency spectrum Ω, which determines the functions $e^{i\omega x}$ that the quantum model "has access to." The second property is the *expressivity of the coefficients* $\{c_\omega\}$ that a class of quantum models can control, which determines how the accessible functions can be combined. In many natural settings, the frequencies are integers, $\Omega \subset \mathbb{Z}^N$, and the sum becomes a multidimensional *partial Fourier series*

$$f_\theta(x) = \sum_{n \in \Omega} c_n(\theta) e^{inx}, \tag{2}$$

where the e^{inx} are orthogonal basis functions. We use the nomenclature *partial* Fourier series to indicate the fact that only a subset of the Fourier coefficients is nonzero. The Fourier series formalism allows us to study quantum models using the rich techniques developed in Fourier analysis.

First, we consider the popular strategy of encoding an input into single-qubit rotations, and show that repeating the encoding r times either sequentially or in parallel allows the model to access frequency spectra Ω consisting of r frequencies.
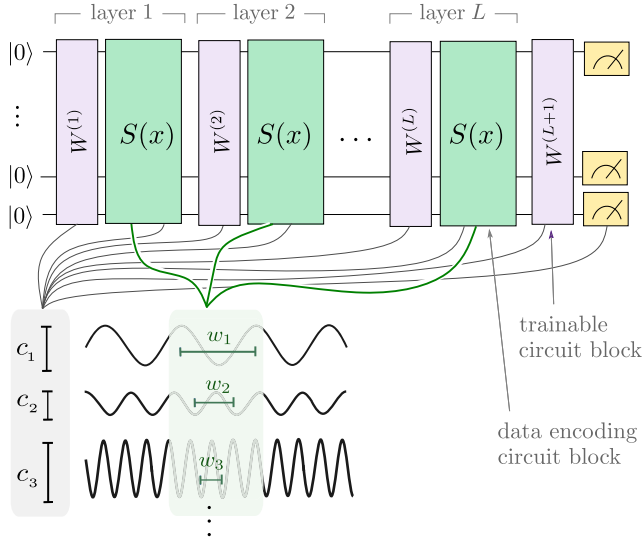
FIG. 1. Illustration of the main result of this paper, shown for one-dimensional inputs $x \in \mathbb{R}$: Quantum models consisting of layers of trainable circuit blocks $W = W(\boldsymbol{\theta})$ and data-encoding circuit blocks $S(x)$ can be written as a weighed sum $\sum_{\omega} c_{\omega} e^{i\omega x}$. The data-encoding circuit determines the frequencies ω, and the remainder of the circuit architecture determines the coefficients c_{ω}. If the ω are integer valued (or integer-valued multiples of a base frequency ω_0), the sum becomes a partial Fourier series, which allows us to systematically study properties of the function class a given quantum model can learn.

This places into a broader context an observation made in Ref. [7], which states that encoding a data feature only once into the angle of a single-qubit rotation restricts the function class that quantum models can learn to a simple sine function (or equivalently, a Fourier series with a single frequency). Second, we provide bounds for the maximum number of frequencies and Fourier coefficients a quantum model can control for more general data-encoding strategies. Finally, we study the *universality* of quantum models. We show that for sufficiently flexible trainable circuit blocks there exists a quantum model which can realize any possible set of Fourier coefficients. If, asymptotically, the accessible frequency spectrum is rich enough, then such models are universal function approximators. This follows from the fact that Fourier series with arbitrary coefficients can approximate any square integrable function on a given interval [16].

A few existing studies are related to our work. For example, Pérez-Salinas *et al.* [17] considered quantum models with sequentially repeated data encodings and conjectured that they are universal function approximators under a special kind of classical data pre-processing. Killoran *et al.* [18] have shown that many neural networks can be naturally emulated on a photonic quantum computer, and point out that such quantum models therefore inherit universality. The majority of quantum-machine-learning papers concerned with questions of expressivity and universality [19–22], however, interpret these concepts from a quantum information perspective, which asks whether a circuit can express *any quantum computation*, not any function in the inputs. However, in the context of (supervised) machine learning, quantum

universality does not necessarily imply universal function approximation; a quantum circuit able to realize arbitrary unitary evolutions may only be able to express a limited class of functions $f(x)$.[1] From the *function* expressivity viewpoint, Ref. [23] has investigated the pseudodimension of a particular class of quantum models, an expressivity metric which allows one to characterize learnability and generalization power of the associated model. Also, the essential role of data encoding for quantum machine learning has been emphasized in previous papers. For example, it was remarked that data encoding determines the features that quantum models represent [24,25], the decision boundaries they can learn [26], as well as the measurements that optimally distinguish between data classes [27]. A central contribution of this paper is to systematically combine the study of data encoding with that of the expressivity of quantum models.

We present our results as follows: Section II introduces the basic idea of writing quantum models as partial Fourier series. Section III puts the tool to use and analyzes the expressivity of quantum models, which leads to a proof that quantum models are universal in Sec. IV. Section V discusses practically relevant implications.

II. QUANTUM MODELS AS PARTIAL FOURIER SERIES

First, we introduce our basic tool: the natural representation of a quantum model as a partial Fourier series. For simplicity, the majority of our presentation will focus on the case of univariate functions with inputs $x \in \mathbb{R}$, but we generalize this to multivariate functions in Appendix A, which is used for the analysis of universality in Sec. IV.

We define a (univariate) quantum model $f_{\boldsymbol{\theta}}(x)$ as the expectation value of some observable with respect to a state prepared via a parametrized quantum circuit, i.e.,

$$f_{\boldsymbol{\theta}}(x) = \langle 0| U^{\dagger}(x, \boldsymbol{\theta}) M U(x, \boldsymbol{\theta}) |0\rangle, \qquad (3)$$

where $|0\rangle$ is some initial state of the quantum computer, $U(x, \boldsymbol{\theta})$ is a quantum circuit that depends on the input x and a (possibly empty) set of parameters $\boldsymbol{\theta}$, and M is some observable. The prediction of the quantum model at a specific point x is estimated in practice by running the circuit multiple times and averaging over the measurement results.[2] The quantum circuit itself is constructed from *L layers*, each consisting of a data-encoding circuit block $S(x)$ and a trainable circuit block $W(\boldsymbol{\theta})$ controlled by the parameters $\boldsymbol{\theta}$ (see Fig. 1). The data-encoding block is the same in every layer and consists of gates of the form $\mathcal{G}(x) = e^{-ixH}$, where H is a Hamiltonian that *generates* the "time evolution" used to encode the data. Since we want to focus on the role of the data encoding, and to avoid further assumptions on how the trainable circuit blocks are

---

[1]As an extreme example, consider a parametrized quantum circuit that encodes the data into gates acting on qubits which are never entangled with the measured qubits—in which case $f(x)$ is a constant function, and the resulting machine learning model trivial.

[2]Note that the quantum model is a theoretical construction, since physical measurements will always result in an *estimate* of the output expectation, making f a random variable—a complication that we will ignore here.
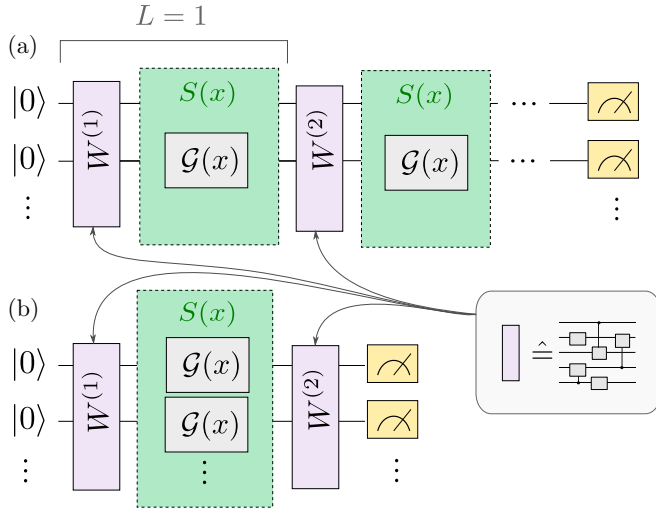
FIG. 2. The general quantum model considered in this paper includes qubit-based circuits where the encoding subroutine consists of a single-qubit gate $\mathcal{G}(x)$, which is often used in practice. The picture illustrates two special cases investigated in Sec. III: (a) shows a circuit where the scalar input feature x is encoded by one single-qubit gate, which can be repeated $r = L > 1$ times but always acts on the same qubit, and (b) repeats the encoding gate r times in "parallel" using only one layer. Note that the trainable blocks W (purple rectangles) represent arbitrary unitaries, which in practice would be implemented as a sequence of local gates (inset).

parametrized, we view the trainable circuit blocks as arbitrary unitary operations, $W(\boldsymbol{\theta}) = W$, and drop the subscript of $f_{\boldsymbol{\theta}}$ from here on.[3] With this assumption, the overall quantum circuit has the form

$$U(x) = W^{(L+1)}S(x)W^{(L)} \cdots W^{(2)}S(x)W^{(1)}. \quad (4)$$

Note that the encoding strategy is very natural, since the physical control parameters of quantum dynamics usually enter as time evolutions of Hamiltonians—the most prominent example being Pauli rotations. This model includes "parallel encodings" that repeat the encoding on different subsystems [28], as well as "data reuploading," where the encoding is repeated multiple times in sequence [17] (see Fig. 2). With a small amount of classical preprocessing this model includes even many quantum-machine-learning algorithms that are not based on the principles of parametrized circuits (see also Sec. V A).

Our goal is to write f as a partial Fourier series

$$f(x) = \sum_{n \in \Omega} c_n e^{inx}, \quad (5)$$

with integer-valued frequencies [if $\Omega = \{-K, \ldots, K\}$, then we call (5) a *truncated Fourier series*]. The first step is to note that one can always find an eigenvalue decomposition of

_____

[3]Of course, in realistic near-term settings these unitaries are implemented as short gate sequences and are by no means universal, and there are many interesting questions around how a specific parametrization influences the properties of the resulting quantum model.

the generator Hamiltonian $H = V^{\dagger} \Sigma V$ where Σ is a diagonal operator containing H's eigenvalues $\lambda_1, \ldots, \lambda_d$ on its diagonal. The data-encoding unitary becomes $S(x) = V^{\dagger} e^{-ix\Sigma} V$, and we can "absorb" $V, V^{\dagger}$ into the arbitrary unitaries $W' = VWV^{\dagger}$. Hence, without loss of generality we will assume that H is diagonal. This allows us to separate the data-dependent expressions from the remainder of the circuit in each component i of the quantum state $U(x)|0\rangle$,

$$[U(x)|0\rangle]_i = \sum_{j_1 \cdots j_L = 1}^{d} e^{-i(\lambda_{j_1} + \cdots + \lambda_{j_L})x} W^{(L+1)}_{ij_L} \cdots W^{(2)}_{j_2 j_1} W^{(1)}_{j_1 1}. \quad (6)$$

For ease of notation we introduce the multi-index $\boldsymbol{j} = \{j_1, \ldots, j_L\} \in [d]^L$, where $[d]^L$ denotes the set of any L integers between $1, \ldots, d$. We can then denote the sum of eigenvalues for a given $\boldsymbol{j}$ by $\Lambda_{\boldsymbol{j}} = \lambda_{j_1} + \cdots + \lambda_{j_L}$, and write

$$[U(x)|0\rangle]_i = \sum_{\boldsymbol{j} \in [d]^L} e^{-i\Lambda_{\boldsymbol{j}} x} W^{(L+1)}_{ij_L} \cdots W^{(2)}_{j_2 j_1} W^{(1)}_{j_1 1}. \quad (7)$$

To consider the full quantum model from Eq. (3) we need to take into account the complex conjugation of this expression as well as the measurement, and get

$$f(x) = \sum_{\boldsymbol{k}, \boldsymbol{j} \in [d]^L} e^{i(\Lambda_{\boldsymbol{k}} - \Lambda_{\boldsymbol{j}})x} a_{\boldsymbol{k}, \boldsymbol{j}}, \quad (8)$$

where the $a_{\boldsymbol{k}, \boldsymbol{j}}$ contain the terms stemming from the arbitrary unitaries and measurement,

$$a_{\boldsymbol{k}, \boldsymbol{j}} = \sum_{i, i'} W^{* \, (1)}_{1 k_1} W^{* \, (2)}_{j_1 j_2} \cdots W^{* \, (L+1)}_{j_L i} M_{i, i'}$$
$$\times W^{(L+1)}_{i' j_L} \cdots W^{(2)}_{j_2 j_1} W^{(1)}_{j_1 1}. \quad (9)$$

The second step consists of grouping all terms in the sum (8) whose basis function $e^{i(\Lambda_{\boldsymbol{k}} - \Lambda_{\boldsymbol{j}})x}$ has the same frequency $\omega = \Lambda_{\boldsymbol{k}} - \Lambda_{\boldsymbol{j}}$. All frequencies accessible to the quantum model are contained in its frequency spectrum

$$\Omega = \{\Lambda_{\boldsymbol{k}} - \Lambda_{\boldsymbol{j}}, \, \boldsymbol{k}, \boldsymbol{j} \in [d]^L\}. \quad (10)$$

This yields

$$f(x) = \sum_{\omega \in \Omega} c_{\omega} e^{i\omega x}, \quad (11)$$

where the coefficients are obtained by summing over all $a_{\boldsymbol{k}, \boldsymbol{j}}$ contributing to the same frequency

$$c_{\omega} = \sum_{\substack{\boldsymbol{k}, \boldsymbol{j} \in [d]^L \\ \Lambda_{\boldsymbol{k}} - \Lambda_{\boldsymbol{j}} = \omega}} a_{\boldsymbol{k}, \boldsymbol{j}}. \quad (12)$$

We note that the frequency spectrum Ω has the following important properties: $0 \in \Omega$, and for every frequency $\omega \in \Omega$, we have also that $-\omega \in \Omega$. Additionally, since $c_{\omega} = c^*_{-\omega}$, Eq. (11) realizes a real-valued function. We will therefore denote with $K = (|\Omega| - 1)/2$ the *size* of the spectrum, as it quantifies how many independent nonzero frequencies the model has access to. The largest available frequency $D = \max(\Omega)$ is called the *degree* of the spectrum. Furthermore, the coefficients c_{ω} are determined by the arbitrary gates $W^{(1)} \cdots W^{(L+1)}$ (which absorbed the $V, V^{\dagger}$ from the encoding Hamiltonians), as well as by the measurement observable.

As a consequence, a quantum model's frequency spectrum is solely determined by the eigenvalues of the data-encoding gates, while its Fourier coefficients depend on the entire circuit, as was claimed in Fig. 1. While so far we have not imposed restrictions on the frequencies ω, one can see that for integer-valued eigenvalues $\lambda_1, \ldots, \lambda_d$, the frequencies in Ω are themselves integer valued, and Eq. (11) yields the real-valued partial Fourier series from Eq. (5). As we will show in the following section, common data-encoding strategies in near-term quantum machine learning fulfill the property of an integer-valued frequency spectrum. Even if the eigenvalues of the encoding gate generators, and therefore the accessible frequencies $\omega \in \Omega$, are merely integer-valued multiples of a "base frequency" $\{n_1\omega_0, n_2\omega_0, \ldots\}$, the treatment is still analogous to the integer case (see Appendix B). We therefore focus much of our analysis on this case.

For both integer or noninteger frequencies, the expressivity of a quantum model is determined by two different properties: the frequency spectrum of the quantum model, including its size and degree, and the expressivity of the coefficients controlled by the model. As we will show in the next section, these two properties give us insights into the function classes that different quantum models can learn.

III. THE EXPRESSIVITY OF QUANTUM MODELS

We proceed to use the Fourier series formalism to investigate the expressivity of quantum models. We start with an analysis of the popular strategy [3,25,29–34] of using single-qubit Pauli rotations in the encoding subroutine $S(x)$ in order to showcase the practical value of the approach. We then characterize the limits of a quantum model's expressivity for a given data-encoding gate in more general terms.

A. A single Pauli-rotation encoding can only learn a sine function

As a "warm-up" application of the Fourier series formalism, we start by considering a simple quantum model with $L = 1$, where we use a single-qubit gate $\mathcal{G}(x) = e^{-ixH}$ to encode the input x into the circuit [see also Fig. 2(a) with $L = 1$],

$$U(x) = W^{(2)}\mathcal{G}(x)W^{(1)}. \tag{13}$$

As a single-qubit gate generator, H has two distinct eigenvalues (λ_1, λ_2). We can without loss of generality always rescale the energy spectrum to $(-\gamma, \gamma)$ because the global phase is unobservable. We note that the class of such encoding gates includes Pauli rotations, with $H = (1/2)\sigma$ for $\sigma \in \{\sigma_x, \sigma_y, \sigma_z\}$, for which $\gamma = \frac{1}{2}$. We aim to show that models of the type (13) *always* lead to functions of the form $f(x) = A\sin(2\gamma x + B) + C$, where A, B, C are constants determined by the nonencoding part of the variational circuit, which reproduces the prior observation from Ref. [7]. A sine function can be described by a truncated Fourier series of degree 1—and in the next section we will go on to show how one can systematically increase the degree by repeating the encoding gate.

First, since we can absorb the factor γ into the data input by rescaling it via $\tilde{x} = \gamma x$, we can assume without loss of

generality that the eigenvalues of H are always $\lambda_1 = -1$, $\lambda_2 = 1$. From Eq. (10) we can immediately see that the spectrum of the quantum model is given by $\Omega = \{-2, 0, 2\}$ [since the possible differences $\lambda_{k_1} - \lambda_{j_1}$ for $\lambda_{k_1}, \lambda_{j_1} \in \{-1, 1\}$ are $-1 - (1)$, $-1 - (-1)$, $1 - (1)$, and $1 - (-1)$]. The Fourier coefficients in Eq. (12) become

$$c_0 = \sum_{i,i'} M_{ii'}(W^*)^{(1)}_{12}(W^*)^{(2)}_{2i}W^{(2)}_{i'1}W^{(1)}_{11}, \tag{14}$$

$$c_2 = \sum_{i,i'} M_{ii'}(W^*)^{(1)}_{11}(W^*)^{(2)}_{1i}W^{(2)}_{i'2}W^{(1)}_{21}, \tag{15}$$

$$c_{-2} = c_2^*, \tag{16}$$

and the quantum model's frequency spectrum consists of a single nonzero frequency,

$$\begin{aligned} f(x) &= c_{-2}e^{i2\tilde{x}} + c_0 + c_2e^{-i2\tilde{x}} \\ &= c_0 + 2|c_2|\cos[2\tilde{x} - \arg(c_2)], \end{aligned}$$

where $\arg(c_2)$ is the complex phase of c_2. For Pauli rotations, one has $\tilde{x} = \gamma x = \frac{x}{2}$, and we recover the result of Ref. [7] with $A = 2|c_2|$, $B = -\pi/2 - \arg(c_2)$, and $C = c_0$. Importantly, we have not assumed anything about the number of qubits, the nature of the unitaries W, or the measurement M. This illustrates a key point of this paper: Even with the ability to implement very wide and deep quantum circuits (which may even be classically intractable to simulate), *the expressivity of the corresponding quantum model is fundamentally limited by the data-encoding strategy*.

To support this finding, Fig. 3 shows numerical evidence: Encoding data via a Pauli-X rotation results in a quantum model that can only learn to fit a Fourier series of a single frequency—and only if that frequency is exactly matched by how the data are scaled.

B. Repeated Pauli encodings linearly extend the frequency spectrum

Given the severe limitations exposed in the previous section, a natural question is how we can extend the accessible frequency spectrum of a quantum model. To this end, we demonstrate in this section that by using either single-layer models with $L = 1$ where the encoding gate is repeated r times in *parallel* [as per Fig. 2(b)], or multilayer models with $L > 1$ where the encoding gate is effectively repeated $r = L$ times in *series* [as per Fig. 2(a)], one can systematically increase the degree of the truncated Fourier series to r. We note once again that both of these techniques have been utilized in prior practical applications [3,17,29–32], and as such the observations we make here offer insight into the properties of these models.

First, let us consider the case of single-qubit Pauli rotations repeated in parallel [Fig. 2(b)]. This is a special case of our base model in Eq. (3), with $L = 1$, and

$$S(x) = e^{-i\frac{x}{2}\sigma_r} \otimes \cdots \otimes e^{-i\frac{x}{2}\sigma_1} \tag{17}$$

$$:= e^{-ixH}, \tag{18}$$

where $\sigma_j \in \{\sigma_x, \sigma_y, \sigma_z\}$. The fact that all rotation gates commute (as they act on different qubits) allows us to diagonalize H by diagonalizing each rotation gate individually. Doing this,
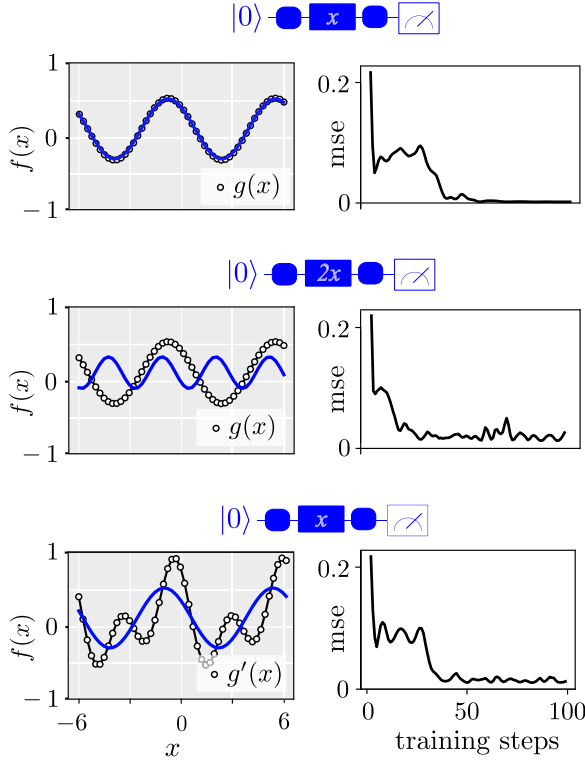
FIG. 3. A parametrized quantum model is trained with data samples (white circles) to fit a target function $g(x) = \sum_{n=-1}^{1} c_n e^{-nix}$ or $g'(x) = \sum_{n=-2}^{2} c_n e^{-nix}$ with coefficients $c_0 = 0.1$, $c_1 = c_2 = 0.15 - 0.15i$. The variational circuit is of the form $f(x) = \langle 0|U^\dagger(x)\sigma_z U(x)|0\rangle$ where $|0\rangle$ is a single qubit, and $U = W^{(2)}R_x(x)W^{(1)}$. The W (round blue symbols) are implemented as general rotation gates parametrized by three learnable weights each, and R_x (square blue symbols) is a single Pauli-X rotation. The left panels show the quantum model function $f(x)$ and target function $g(x)$, $g'(x)$, while the right panels show the mean-squared error between the data sampled from g and f during a typical training run. Feeding in the input x as is (top row), the quantum model easily fits the target of degree 1. Rescaling the inputs $x \to 2x$ causes a frequency mismatch, and the model can no longer learn the target (middle row). However, even with the correct scaling, the variational circuit cannot fit the target function of degree 2 (bottom row). The experiments in this paper were all performed using the PennyLane software library [35].

we find that

$$S(x) = V_r e^{-i\frac{x}{2}\sigma_z}V_r^\dagger \otimes \cdots \otimes V_1 e^{-i\frac{x}{2}\sigma_z}V_1^\dagger \tag{19}$$

$$= V \exp\left(-i\frac{x}{2}\sum_{q=1}^{r}\sigma_z^{(q)}\right)V^\dagger \tag{20}$$

$$:= V e^{-ix\Sigma}V^\dagger, \tag{21}$$

where $\sigma_z^{(q)}$ is the (diagonal) r-qubit operator which acts nontrivially, via σ_z, only on the qth qubit. Performing the calculation yields $\Sigma = \text{diag}(\lambda_1, \ldots, \lambda_{2^r})$, with the $r + 1$ unique entries

$$\lambda_p = \left(\frac{p}{2} - \frac{r-p}{2}\right) = p - \frac{r}{2}, \quad p \in \{0, \ldots, r\},$$

which are all possible sums of r values $\pm 1/2$. According to Eq. (10), the frequency spectrum for $L = 1$ contains differences of any two of these eigenvalues, and we get

$$\Omega_{\text{par}} = \left\{\lambda_{k_1} - \lambda_{j_1}\,\middle|\, k_1, j_1 \in \{1, \ldots, 2^r\}\right\} \tag{22}$$

$$= \left\{\left(p - \frac{r}{2}\right) - \left(p' - \frac{r}{2}\right)\middle| p, p' \in \{0, \ldots, r\}\right\} \tag{23}$$

$$= \{p - p' \mid p, p' \in \{0, \ldots, r\}\} \tag{24}$$

$$= \{-r, -(r-1), \ldots, 0, \ldots, r-1, r\}. \tag{25}$$

Hence, a univariate quantum model with r parallel Pauli-rotation encodings can be expressed as a truncated Fourier series of degree r.

Interestingly, the same scaling effect is achieved by a single-qubit Pauli-rotation encoding repeated layerwise [Fig. 2(a)]. Consider the quantum model in Eqs. (3) and (4), for $L = r > 1$ layers, where $S(x) = \exp[-i(x/2)\sigma_j]$ is a single-qubit Pauli rotation (i.e., $\sigma_j \in \{\sigma_x, \sigma_y, \sigma_z\}$) which acts on the same qubit in each layer. The circuit in Eq. (4) becomes

$$U(x) = W^{(L+1)}e^{-i\frac{x}{2}\sigma_L}W^{(L)} \cdots W^{(2)}e^{-i\frac{x}{2}\sigma_1}W^{(1)}.$$

Diagonalizing the Pauli rotations as before, then gives us $\Sigma = (1/2)\sigma_z$ for all encoding layers. The frequency spectrum from Eq. (10) is a sum of $2r$ terms of value $\pm 1/2$,

$$\Omega_{\text{seq}} = \left\{\left(\lambda_{k_1} + \cdots + \lambda_{k_r}\right) - \left(\lambda_{j_1} + \cdots + \lambda_{j_r}\right)\middle|\right.$$
$$\left. k_1, \ldots, k_r, j_1, \ldots, j_r \in \{1, 2\}\right\}. \tag{26}$$

After a short calculation, one finds that $\Omega_{\text{seq}} = \Omega_{\text{par}}$. Again, a quantum model with r sequential repetitions of the single-qubit Pauli encoding can be expressed as a truncated Fourier series of degree r. The growth mechanism of a quantum model's frequency spectrum via parallel and sequential repetitions of single-qubit Pauli encodings is numerically illustrated in Fig. 4.

C. Limits of expressivity

The representation of quantum models as Fourier-type sums immediately allows us to derive upper bounds on the expressivity of such quantum models when using L repetitions of an encoding gate of dimension d (which is at most the size of the overall Hilbert space). First, let us consider the maximum spectrum size $K(L, d)$ of a quantum model, quantifying the number of frequencies it can "support" or "has access to." Since the frequency spectrum is defined as $\Omega = \{(\lambda_{j_1} + \cdots + \lambda_{k_L}) - (\lambda_{j_1} + \cdots + \lambda_{k_L})\}$ (where the indices $j_1, \ldots, j_L$, $k_1, \ldots, k_L$ run over all dimensions of the encoding gate, from 1 to d), the frequencies are sums of $2L$ terms, each having d potential values. As a result, they can at most realize d^{2L} distinct values—irrespective of whether the eigenvalues are real or integer valued. Since the size K counts the pairs $-\omega, \omega \in \Omega$ as one and excludes the "zero frequency," we get

$$K \leqslant \frac{d^{2L}}{2} - 1. \tag{27}$$

As an example, if data are encoded in a single-qubit encoding gate, we recover the result from the previous sections where the model has degree $\frac{2^2}{2} - 1 = 1$. Using L different encoding gates increases this to $\frac{2^{2L}}{2} - 1$. As we have seen, further
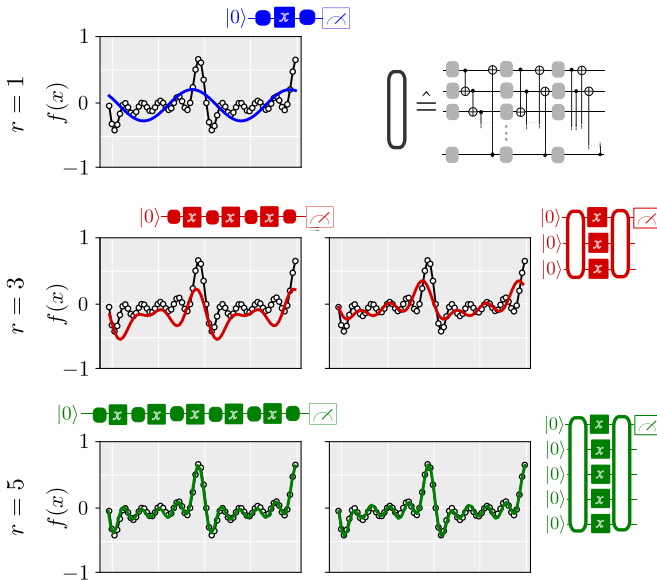
FIG. 4. Fitting a truncated Fourier series of degree 5, $g(x) = \sum_{n=-5}^{5} c_n e^{2inx}$ with $c_n = 0.05 - 0.05i$ for $n = 1, \ldots, 5$ and $c_0 = 0$, using a quantum model that repeats the encoding $r = 1, 3, 5$ times in sequence (left) and in parallel (right). Increasing r allows for closer and closer fits until $r = 5$ fits the data almost perfectly in both cases—illustrating that parallel and sequential repetitions of Pauli encodings extend the Fourier spectrum in the same manner. All models were trained with at most 200 steps of an Adam optimizer with learning rate 0.3 and batch size 25. For the "parallel" simulations, the W are not arbitrary unitaries but implemented by a smaller ansatz of three layers of parametrized rotations as well as entangling controlled-NOT (CNOT) gates, as per Ref. [29], which is depicted by the open rounded gate symbols. The quantum model still easily fitted the target function, which suggests that the results of this paper are of relevance for realistic quantum models.

assumptions on the eigenvalues allow us to make this bound a lot tighter, for example, when the L repetitions use *the same* single-qubit encoding gate, $K = L$.

An interesting question is whether there is a single quantum gate which can encode data into a quantum model that supports the frequency spectrum $\Omega_\infty = \{-\infty, \ldots, -1, 0, 1, \ldots, \infty\}$ of a full Fourier series. The answer is yes: The ubiquitous phase shifts in continuous-variable (CV) quantum systems, which correspond to a free evolution of a harmonic oscillator, have the number operator $\hat{n} = \text{diag}(0, 1, 2, \ldots)$ as a generator.

While the frequency spectrum of a quantum model can be directly derived from the input encoding gates, the flexibility in the coefficients is a lot harder to investigate systematically (we will do so for special cases in the universality proofs in Sec. IV). In principle, every block $W^{(1)}, \ldots, W^{(L+1)}$, as well as the measurement observable, contribute to every Fourier coefficient. This means that only a few degrees of freedom in the gates may change an exponentially large (or, in the case of continuous-variable quantum computing, infinite) amount of Fourier coefficients. However, these Fourier coefficients are not arbitrary, but functions of the limited degrees of freedom of the quantum circuit, and a quantum circuit of a certain structure may only be able to realize a small subset of the

entire set of all possible Fourier coefficients $\{c_n\}$. To arbitrarily control $K + 1$ complex Fourier coefficients, we need *at least* $M \geqslant 2K + 1$ real degrees of freedom—in other words, parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$—in the quantum circuit. As a special case, we saw that repeating a Pauli encoding L times supports a spectrum of size L, which means that we need at least $2L$ degrees of freedom in the quantum circuit to control the Fourier coefficients arbitrarily—a scaling that is realistic for shallow circuits to "utilize the full power" of the frequency spectrum.

While a systematic analysis of how a parametrized ansatz for the trainable blocks W impacts the control of a quantum model's Fourier coefficients exceeds the scope of this paper, our simulations suggest that even quantum models with shallow trainable circuit blocks W give rise to rich subsets of Fourier coefficients (see Fig. 5). However, as the figure shows, an ansatz may structurally set a certain Fourier coefficient to zero. An interesting further observation is that the variance of the coefficients decreases with higher orders. Mathematically, this property stems from the fact that the number of terms in the sum of Eq. (12) tends to decrease with larger frequencies, since there are fewer ways to construct those frequencies by the difference $\Lambda_j - \Lambda_k$ of sums of encoding generator eigenvalues. We note that the Fourier coefficients of square-integrable functions show a similar behavior, which contributes to the convergence of such series.

IV. QUANTUM MODELS ARE ASYMPTOTICALLY UNIVERSAL

In the previous sections we have seen, at least for univariate functions, that certain quantum models can be written as a partial Fourier series, in which the accessible frequencies are fully determined by the spectra of the Hamiltonians generating the data-encoding gates. Additionally, by using Pauli rotations as an explicit example, we have shown that by repeating such encodings, either in parallel ($L = 1$) or in a series ($L > 1$), it is possible to realize a *truncated* Fourier series, with the number of accessible frequencies determined by the number of data-encoding gate repetitions. In light of these results, it is clear that if we allow for sufficiently many repetitions of simple data-encoding gates (such as Pauli rotations), or for Hamiltonians with large enough dimension and suitably nondegenerate spectra, then quantum models can realize arbitrary frequency spectra.

However, as discussed in the previous section, the expressivity of a quantum model is determined not only by the accessible frequency spectrum, but also by the flexibility one has in adjusting the contributions of the frequencies, i.e., with which flexibility the Fourier coefficients can be chosen. In this section we show that if one allows for trainable circuit blocks which are flexible enough to realize arbitrary global unitaries, then there exists an $L = 1$ quantum model which can realize all possible sets of Fourier coefficients. Combined with the observations from the previous sections, this allows us to show that such quantum models are asymptotically universal, in the sense that if we allow the global Hilbert space dimension (or the number of finite dimensional subsystems) to tend to infinity, then such a quantum model can approximate, to arbitrary accuracy, any square-integrable function on a suitable domain.
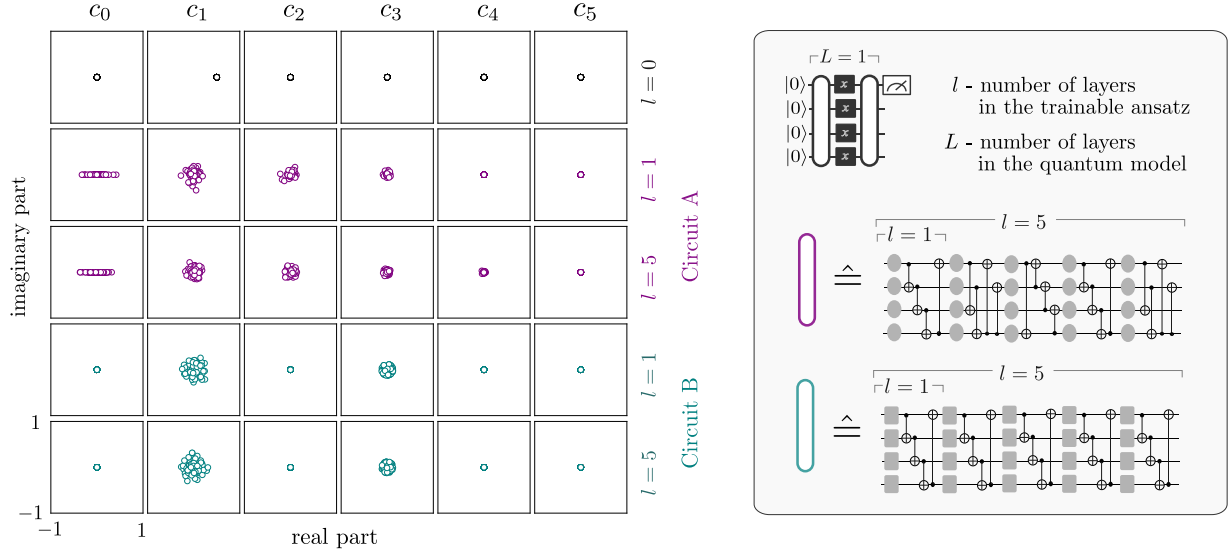
FIG. 5. Real and imaginary parts of the first six Fourier coefficients sampled from 100 randomly initialized $L = 1$ quantum models. The models share the same encoding strategy of parallel Pauli-X rotations (square symbols) but vary in the ansatz and number of layers for the trainable unitaries W. Circuit A uses an ansatz of trainable arbitrary single-qubit rotations and layer-dependent entangling structure proposed in Ref. [29] and already used in Fig. 4, while circuit B uses trainable Pauli-X rotations with a simple entangling structure. The plots suggest that the "expressivity" of the trainable circuit block—here represented by increasing the number of times l an ansatz is repeated—has little influence on the distribution of the Fourier coefficients, as opposed to the type of ansatz.

More specifically, we consider the (multivariate) single-layer quantum model $f_{\boldsymbol{\theta}} : \mathbb{R}^N \to \mathbb{R}$ defined via

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \langle 0 | U^\dagger(\boldsymbol{\theta}, \boldsymbol{x}) M U(\boldsymbol{\theta}, \boldsymbol{x}) | 0 \rangle, \qquad (28)$$

where

$$U(\boldsymbol{\theta}, \boldsymbol{x}) = W^{(2)}(\boldsymbol{\theta}^{(2)}) S(\boldsymbol{x}) W^{(1)}(\boldsymbol{\theta}^{(1)}), \qquad (29)$$

with $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)} \subseteq \boldsymbol{\theta}$ and

$$S(\boldsymbol{x}) := e^{-ix_1 H_1} \otimes \cdots \otimes e^{-ix_N H_N}. \qquad (30)$$

The above model is a natural extension of the univariate $L = 1$ model we explored in previous sections. In Appendix A we show that it naturally realizes a multivariate Fourier series, with the frequency spectrum fully determined by the spectra of the data-encoding Hamiltonians $\{H_l\}$, and the Fourier coefficients determined by the remainder of the circuit.

It is important to emphasize that in practical applications one would typically consider trainable circuit blocks whose circuit depth scales in a controlled way with respect to the number of qubits in the circuit. However, we will in this work assume that the trainable circuit blocks are sufficiently flexible to realize arbitrary global unitaries, which may require exponential circuit depth when decomposed into natural primitive gate sets. Given this, the asymptotic universality of quantum models with either constant, logarithmic, or polynomial circuit depth trainable blocks remains an interesting open question.

With this assumption on the trainable circuit blocks, we can drop the explicit dependence on $\boldsymbol{\theta}$, and by absorbing $W^{(1)}$ into the initial state $|\Gamma\rangle$, and $W^{(2)}$ into the observable M, consider instead the equivalent model

$$f(\boldsymbol{x}) = \langle \Gamma | S^\dagger(\boldsymbol{x}) M S(\boldsymbol{x}) | \Gamma \rangle, \qquad (31)$$

where the universality of $W^{(1)}$ and $W^{(2)}$ is reinterpreted as the assumption that $|\Gamma\rangle$ can be an arbitrary state, and M an arbitrary observable. In order to simplify things further, we will also make the additional assumption that all data-encoding Hamiltonians are equal, i.e., that

$$S(\boldsymbol{x}) := e^{-ix_1 H} \otimes \cdots \otimes e^{-ix_N H} \qquad (32)$$

$$:= S_H(\boldsymbol{x}). \qquad (33)$$

We are interested in a reasonable notion of universality in the *asymptotic* regime of infinitely many available subsystems. To formalize this, we introduce the concept of a *Hamiltonian family* $\{H_m \,|\, m \in \mathbb{N}\}$ where H_m acts on m subsystems of dimension d. An explicit example of such a family is a simple tensor product of Pauli rotations, as studied in Sec. III B, which corresponds to the Hamiltonian

$$H_m = \sum_{i=1}^m \sigma_q^{(i)}. \qquad (34)$$

As illustrated in Fig. 6, such a Hamiltonian family defines a family of models $\{f_m\}$ via

$$f_m(\boldsymbol{x}) = \langle \Gamma | S_{H_m}^\dagger(\boldsymbol{x}) M S_{H_m}(\boldsymbol{x}) | \Gamma \rangle, \qquad (35)$$

where for each m, the measurement M and the state $|\Gamma\rangle$ (or equivalently the unitaries $W^{(1)}$ and $W^{(2)}$) are the learnable elements of the model.

Now, given some Hamiltonian H_m with eigenvalues $\{\lambda_1, \ldots, \lambda_{d^m}\}$, we call

$$\Omega_{H_m} = \{\lambda_j - \lambda_k \,|\, j, k \in \{1, \ldots, d^m\} \qquad (36)$$

the *frequency spectrum* associated with H_m. To achieve universality, we need a Hamiltonian family whose frequency spectrum asymptotically contains any integer frequency.
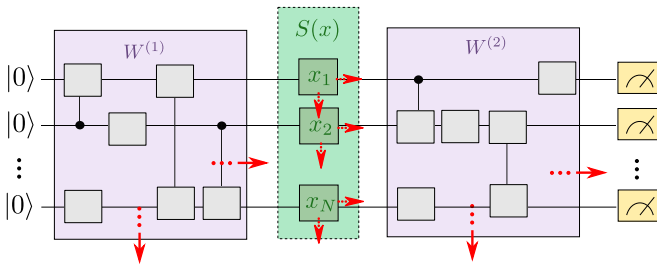
FIG. 6. Multivariate $L = 1$ quantum model considered for the universality theorem. Here, $S(x)$ consists of feature-encoding gates acting on different subsystems (green boxes). The Hamiltonians that generate these gates are defined to increase the "richness" of their spectrum with growing dimension of the subsystems (red arrows). Since we assume that the circuit depth and structure of the trainable unitaries $W^{(1)}$ and $W^{(2)}$ is sufficient to allow for the realization of arbitrary unitary operations, the trainable circuits grow in dimension along with the total system size.

We formalize this via the following notion: A Hamiltonian family $\{H_m\}$ is a *universal Hamiltonian family* if it has the property that for all $K \in \mathbb{N}$ there exists some $m \in \mathbb{N}$ such that

$$\mathbb{Z}_K = \{-K, \ldots, 0, \ldots, K\} \subseteq \Omega_{H_m}. \tag{37}$$

As we have seen in the previous section, the Hamiltonian family defined by the Hamiltonians in Eq. (34) is indeed a universal Hamiltonian family, with $m = K$. As the possible number of frequencies grows exponentially, one could think of more complicated Hamiltonian families in which the required number of available subsystems only grows logarithmically $m = O(\log K)$, at the cost of more complicated, global Hamiltonian terms. With this setup, we can now state the following universality result:

Theorem. Let $\{H_m\}$ be a universal Hamiltonian family, and $\{f_m\}$ the associated quantum model family, defined via Eq. (35). For all functions $g \in L_2([0, 2\pi]^N)$, and for all $\epsilon > 0$, there exists some $m' \in \mathbb{N}$, some state $|\Gamma\rangle \in \mathbb{C}^{d^{m'}}$, and some observable M such that

$$||f_{m'} - g||_2 \leqslant \epsilon. \tag{38}$$

A full proof is given in Appendix C, however, in the following we will provide a sketch of the proof in order to give an outline of the ideas and techniques. The proof begins by noting that any square-integrable function g on a finite interval can be approximated by a truncated Fourier series to arbitrary precision. We therefore reduce the task to finding a quantum model for this truncated Fourier series. The universality property of the Hamiltonian family implies that the multivariate models we consider can express all necessary frequencies to perform that approximation. We then show how to use the freedom in choosing the initial state and the observable to reproduce the truncated Fourier series of g exactly, leading to an approximation by a quantum model with arbitrary precision.

Note that the statement that there exists some state $|\Gamma\rangle$ and some observable M is equivalent to the statement that the target function can be learned by the relevant model (under the assumption the trainable circuit blocks are sufficiently flexible). As any frequency spectrum is asymptotically accessible, due to the assumption of a universal Hamiltonian family, the

universality theorem is essentially equivalent to the statement that with sufficiently flexible circuit blocks such quantum models can realize any set of Fourier *coefficients*.

V. PRACTICAL IMPLICATIONS FOR QUANTUM MACHINE LEARNING

In this last section, we discuss the scope and practical relevance of our results for quantum machine learning. First, we motivate that many quantum models proposed in the literature that do not immediately fit the base model from Eq. (3) can still be analyzed within our framework under the assumption that they encode classically preprocessed features $\phi(x)$ instead of the original features x. Second, we summarize guidelines that can help with the design of quantum-machine-learning algorithms.

A. Classical preprocessing

The base model used in this paper makes the assumption that a data feature is encoded into a subroutine $S(x)$ which consists of gates $\mathcal{G}(x) = e^{-ixH}$. We motivate in this section that many quantum-machine-learning algorithms which use other strategies of data encoding actually perform an implicit preprocessing of the data, and then use the "time-evolution" encoding studied here. The results of this paper are hence valid for the different features resulting from the preprocessing step.

For example, the standard encoding procedure of traditional [36] and some noisy intermediate-scale quantum (NISQ) [4] quantum algorithms, associates the n-bit binary representation of each (scalar) input feature x with an n-qubit basis state, such as $x \mapsto |01011\rangle$. The preprocessing step therefore maps original features x to the angles $\phi(x) = [\phi_1(x), \ldots, \phi_n(x)]$ with which the n qubits have to be rotated to reflect every binary decimal digit of x (i.e., π for $|1\rangle$ or 0 for $|0\rangle$). Our investigation here states that the quantum model for a single input feature corresponds to a multidimensional Fourier series in the angles, with a frequency spectrum size of at most n. In other words, the preprocessing changed the accessible Fourier spectrum by changing the features.

Another example is so-called "amplitude encoding" (i.e., Refs. [28,29]), which associates an input vector x with the values of the amplitudes of a quantum state. Practically, this requires $S(x)$ to be an arbitrary state preparation routine that is parametrized by some angles computed from x. The classical preprocessing therefore maps the original input to the set of angles used in the state preparation, $x \mapsto \phi(x)$.

Preprocessing is also sometimes used in encoding strategies that directly feed input features into Pauli rotations. One example was used in Figs. 3 and 4, where we rescaled the inputs by a classical hyperparameter. In Ref. [17] it has been proposed to make these hyperparameters trainable (which in the light of the present analysis would allow for an adaptive "frequency matching" and may help to increase the expressivity of small quantum circuits). Another example is to construct higher-order features that are arithmetic combinations of the original inputs, such as $\phi_1(x) = x_1 x_2$, $\phi_2(x) = x_2 x_3, \ldots$, as used in the quantum feature map proposed in Ref. [25].

These examples suggest that implicit preprocessing can extend the function classes that quantum models can learn even further. However, care needs to be taken when making theoretical claims about the power of a quantum-machine-learning algorithm, which is, strictly speaking, a result of the quantum algorithm *plus* the specific preprocessing strategy. In particular, comparisons to classical machine learning models should identify the preprocessing strategy and consider feeding the same preprocessed features to the classical model.

B. Practical insights

Finally, we want to summarize how the results of this paper can be used to understand and evaluate different design decisions of quantum-machine-learning models:

(1) If data are encoded via a Hamiltonian time evolution, we can naturally describe the class of functions that quantum models can learn as a partial Fourier series. The Hamiltonian defines the available frequencies in the series, and the gates that do not encode data define the Fourier coefficients.

(2) If data are encoded into single-qubit Pauli rotations, the number of rotations used limits the number of frequencies that the model has access to. Repeating an encoding gate can help to increase the frequency spectrum, and thereby the expressivity of a quantum model.

(3) Quantum models naturally learn periodic functions in the data. One should therefore consider appropriate data rescaling strategies, to make sure the data lie within the period of the function class. The natural representation of quantum models as a Fourier series may suggest that time-series learning and signal processing tasks are particularly suitable applications for quantum machine learning. It may also hint at inherent regularizing properties of quantum models that exclude higher-order Fourier frequencies.

(4) Classical preprocessing of the data, such as creating more features, can give small models more expressivity by enriching the frequency spectrum.

(5) Adjusting the entries of the observable freely was a key ingredient in proving universality of quantum circuits in Sec. IV. Fixing the observable in a quantum model therefore limits its applicability. This fact suggests that parametrizing the observable itself may be a key ingredient for flexible quantum models.

Ideally, one would hope that our results could provide concrete guidelines for the design of quantum-machine-learning models. However, in practical settings the process of model selection should be guided not purely by model expressivity, but rather through the expected generalization performance of the model function class, as captured by capacity metrics such as the Vapnik-Chervonenkis (VC) dimension or Rademacher complexity [37]. While such capacities can be calculated for very simple function classes, calculating such metrics for more complex model classes, such as the quantum models studied here, is significantly harder. Additionally, in modern overparametrized models, which can often fit even randomized training data perfectly [38], more sophisticated approaches are necessary to understand the generalization capacity [39]. In light of this, the insights on how to make models more expressive should not be misinterpreted as recommendations for how to design *good* quantum models—a

question which is much more complex and whose answer depends strongly on the context.

VI. CONCLUSION

In this paper we presented a systematic mapping between a large class of quantum-machine-learning models and partial Fourier series, which has allowed us to explore and quantify the effect of commonly used data-encoding mechanisms on the expressivity of these quantum models. We believe that this framework both lays a foundation for further theoretical analysis, and can serve as a useful guide in the search for suitable applications of such models. Additionally, this work provides a connection between quantum machine learning and ideas from the classical machine learning literature, such as neural networks with periodic activation functions [40], and parametrized Fourier series as an alternative to neural networks [41,42].

As mentioned throughout the paper, a variety of interesting questions remain. First, can the framework developed here help us to understand and quantify the generalization capacity of quantum models, and therefore guide model selection in a meaningful way? In particular, by using the representation of a quantum model as a partial Fourier series, can one calculate meaningful modern generalization measures [39] and use these for the development of model-selection guidelines? Second, we have proven our universality result under the assumption of exponential depth trainable circuit blocks (which provides a reasonable notion of *asymptotic* universality with respect to circuit depth). In practical settings, however, one is interested in trainable circuit blocks with depth restrictions. Can one prove universality of such quantum models with either constant, logarithmic, or polynomial depth trainable circuit blocks? In order to answer this question our toolbox needs to be developed further to understand how the structure of the trainable circuit blocks influences the set of accessible Fourier coefficients. Finally, it is currently unclear for which concrete applications quantum models may be naturally suited, or offer any sort of advantage over classical techniques, such as neural networks. Another question is therefore whether one can use knowledge of the function class expressed by quantum models, as developed in this work, to suggest natural applications for quantum machine learning.

Auxiliary information regarding the CO_2 emissions related to this work is supplied in the Supplemental Material [43].

Note added. Recently, we became aware that the connection between Fourier series and quantum-machine-learning models with repeated data encoding has already been established in Ref. [44]. While there is significant overlap between this work and ours, we provide a different universality result, as well as a systematic development of this connection through practically relevant examples.

Code to reproduce the figures and explore further settings can be found in the GitHub repository [45].

APPENDIX A: PARTIAL FOURIER SERIES REPRESENTATION OF MULTIVARIATE FUNCTIONS

In this Appendix we show how a certain class of $L = 1$ quantum models naturally realize a multivariate Fourier series. On the one hand, this shows a way in which the univariate case analyzed in the paper can easily generalize to multivariate models by encoding the features into different quantum subsystems. On the other hand, the multivariate model described in this Appendix is a quantum model whose asymptotic universality is stated and discussed in Sec. IV, and proven in Appendix C.

More specifically, we consider a quantum model of the form

$$f(\boldsymbol{x}) = \langle 0|[W^{(1)\dagger}S^\dagger(\boldsymbol{x})W^{(2)\dagger}]M[W^{(2)}S(\boldsymbol{x})W^{(1)}]|0\rangle, \quad \text{(A1)}$$

where

$$S(\boldsymbol{x}) := e^{-ix_1 H_1} \otimes \cdots \otimes e^{-ix_N H_N}. \quad \text{(A2)}$$

Without loss of generality, instead of explicitly considering arbitrary unitaries $W^{(1)}$ and $W^{(2)}$, we can "absorb" the unitaries into the initial state and measurement and consider the equivalent model

$$f(\boldsymbol{x}) = \langle \Gamma|S^\dagger(\boldsymbol{x})MS(\boldsymbol{x})|\Gamma\rangle, \quad \text{(A3)}$$

where

$$|\Gamma\rangle = \sum_{j_1,\ldots,j_N=1}^{2^d} \gamma_{j_1,\ldots,j_N}|j_1\rangle \otimes \cdots \otimes |j_N\rangle \quad \text{(A4)}$$

is some arbitrary state, and M is some arbitrary observable. To simplify the index handling, we introduce the multi-indices $\boldsymbol{j} \in [2^d]^N$ with which we can rewrite

$$|\Gamma\rangle := \sum_{\boldsymbol{j}} \gamma_{\boldsymbol{j}}|\boldsymbol{j}\rangle. \quad \text{(A5)}$$

Additionally, as argued before, we can without loss of generality assume that all Hamiltonians are diagonal, i.e., that

$$H_k = \text{diag}\big(\lambda_1^{(k)}, \ldots, \lambda_{2^d}^{(k)}\big). \quad \text{(A6)}$$

With this assumption, we note that $S(\boldsymbol{x})$ is diagonal with entries

$$[S(\boldsymbol{x})]_{\boldsymbol{j},\boldsymbol{j}} = e^{-i\boldsymbol{x}\cdot\boldsymbol{\lambda}_{\boldsymbol{j}}}, \quad \text{(A7)}$$

where we have defined

$$\boldsymbol{\lambda}_{\boldsymbol{j}} = \big(\lambda_{j_1}^{(1)}, \ldots, \lambda_{j_N}^{(N)}\big). \quad \text{(A8)}$$

Given this, we see that

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{j}} \sum_{\boldsymbol{k}} \gamma_{\boldsymbol{j}}^* \gamma_{\boldsymbol{k}} [S^\dagger(\boldsymbol{x})MS(\boldsymbol{x})]_{\boldsymbol{j},\boldsymbol{k}} \quad \text{(A9)}$$

$$= \sum_{\boldsymbol{j}} \sum_{\boldsymbol{k}} \gamma_{\boldsymbol{j}}^* \gamma_{\boldsymbol{k}} M_{\boldsymbol{j},\boldsymbol{k}} e^{i\boldsymbol{x}\cdot(\boldsymbol{\lambda}_{\boldsymbol{k}}-\boldsymbol{\lambda}_{\boldsymbol{j}})}, \quad \text{(A10)}$$

which is indeed a partial multivariate Fourier series, with the accessible frequencies fully determined by the spectra of

the encoding Hamiltonians $\{H_k\}$, and the Fourier coefficients determined by the trainable unitaries (or equivalently, the state and observable).

APPENDIX B: NONINTEGER FREQUENCIES

In the main text, we put our focus on quantum models with integer-valued frequency spectra, as they naturally arise when using Pauli-rotation gates and allow for an analysis with the techniques of a Fourier series. Here, we will briefly discuss why many quantum models with non-integer-valued frequency spectra can be treated similarly.

First, note that we can always decompose functions of the form $e^{i\omega x}$ into a Fourier series of integer-valued frequencies, i.e.,

$$e^{i\omega x} = \sum_{n=-\infty}^{\infty} \frac{(-1)^n \sin\omega\pi}{(\omega - n)\pi} e^{inx} = \sum_{n=-\infty}^{\infty} \text{sinc}(\omega - n)e^{inx}, \quad \text{(B1)}$$

with $\text{sinc}(z) = \sin(\pi z)/\pi z$. However, as we can see from this expression, any noninteger frequency in general "contributes" to *infinitely many* Fourier coefficients. It turns out that a rather general case of quantum models with noninteger frequencies can be handled equivalently, namely if the frequencies are integer multiples of some basic frequency ω_0,

$$\Omega = \{0, \pm n_1\omega_0, \pm n_2\omega_0, \ldots\}. \quad \text{(B2)}$$

This condition is equivalent to all frequencies in Ω being mutually commensurable, i.e., the ratio of any two frequencies is a rational number. This is the case in many natural settings, for example if encodings with noninteger frequencies are repeated in parallel alike to the Pauli encodings in Sec. III.

Basis functions of the form $e^{ixn\omega_0}$ are periodic functions on the interval $[0, 2\pi/\omega_0]$. This means that the generated Fourier-type sum in Eq. (11) can be understood as the partial Fourier series in Eq. (5), but on a different interval. Alternatively, one can imagine rescaling the data by $\tilde{x} = x/\omega_0$, with which

$$e^{i\frac{x}{\omega_0}\omega} = e^{i\tilde{x}n}. \quad \text{(B3)}$$

While this strategy could in principle be applied to any frequency spectrum where the frequencies are mutually commensurable, one has to be aware that ω_0 is as least as small as the smallest difference between frequencies in Ω. If very close frequencies are present in the spectrum, the data will have to be rescaled by a very large factor to an interval where the generated Fourier coefficients may be sparse and the approximation quality poor.

APPENDIX C: PROOF OF THE UNIVERSALITY THEOREM

We provide in this Appendix a proof of the universality theorem stated in Sec. IV, which we restate for completeness:

Theorem. Let $\{H_m\}$ be a universal Hamiltonian family, and $\{f_m\}$ the associated quantum model family, defined via Eq. (35). For all functions $g \in L_2([0, 2\pi]^N)$, and for all $\epsilon > 0$, there exists some $m' \in \mathbb{N}$, some state $|\Gamma\rangle \in \mathbb{C}^{d^{m'}}$, and some

observable M such that

$$||f_{m'} - g||_2 \leqslant \epsilon. \tag{C1}$$

Proof. To begin with, we note that we can approximate any given $g \in L_2([0, 2\pi]^N)$, up to an arbitrarily small error in L_2 norm, by using a truncated Fourier series [46]. More specifically, for any given $\epsilon > 0$, there exists some $K \in \mathbb{N}$ and some set of coefficients $\{c_{\boldsymbol{n}} \mid \boldsymbol{n} \in \mathbb{Z}_K^N\}$, with $c_{\boldsymbol{n}} = c_{-\boldsymbol{n}}^*$, such that

$$\tilde{g}(\boldsymbol{x}) = \sum_{n_1 = -K}^{K} \cdots \sum_{n_N = -K}^{K} c_{\boldsymbol{n}} e^{i\boldsymbol{x} \cdot \boldsymbol{n}} \tag{C2}$$

$$:= \sum_{\boldsymbol{n} \in \mathbb{Z}_K^N} c_{\boldsymbol{n}} e^{i\boldsymbol{x} \cdot \boldsymbol{n}} \tag{C3}$$

satisfies

$$||\tilde{g} - g|| \leqslant \epsilon. \tag{C4}$$

In order to prove the theorem we therefore only need to show that there exists an $m' \in \mathbb{N}$, some state $|\Gamma\rangle$, and some observable M so that the associated quantum model $f_{m'}$ generates the Fourier series $\tilde{g}$. Recall that the quantum model was defined as

$$f_m(\boldsymbol{x}) = \langle \Gamma | S_{H_m}^\dagger(\boldsymbol{x}) M S_{H_m}(\boldsymbol{x}) | \Gamma \rangle, \tag{C5}$$

with

$$S_{H_m}(\boldsymbol{x}) := e^{-ix_1 H_m} \otimes \cdots \otimes e^{-ix_N H_m}. \tag{C6}$$

In Appendix A, we have seen that we can express the output of the model as

$$f_m(\boldsymbol{x}) = \sum_j \sum_k \gamma_j^* \gamma_k M_{j,k} e^{i\boldsymbol{x} \cdot (\lambda_k - \lambda_j)}, \tag{C7}$$

where the multi-indices $\boldsymbol{j}$ and $\boldsymbol{k}$ have N entries that iterate over all 2^d basis states of the d qubit subsystems. Let Ω_{H_m} be the frequency spectrum of H_m, as defined in Eq. (36). As the $\{H_m\}$ form a universal family of Hamiltonians by assumption, we can choose an $m' \in \mathbb{N}$ so that

$$\mathbb{Z}_K = \{-K, \ldots, 0, \ldots, K\} \subseteq \Omega_{H_{m'}}. \tag{C8}$$

The accessible frequency vectors $\lambda_j - \lambda_k$ independently contain all possible combinations of the frequencies in $\Omega_{H_{m'}}$. The vector-valued frequency spectrum for the multivariate case is therefore the Cartesian product of N copies of $\Omega_{H_{m'}}$:

$$\Omega = \underbrace{\Omega_{H_{m'}} \times \cdots \times \Omega_{H_{m'}}}_{N \text{ times}}. \tag{C9}$$

As $\mathbb{Z}_K \subseteq \Omega_{H_{m'}}$ we naturally have that $\mathbb{Z}_K^N \subseteq \Omega$, which means that the Fourier series generated by the chosen model contains all terms that are necessary to construct the Fourier series $\tilde{g}$.

We can now revisit Eq. (C7) and show that we can leverage the freedom of choosing both the initial state $|\Gamma\rangle$ and the observable M arbitrarily to adjust all terms in the sum of Eq. (C7) freely up to the complex-conjugation symmetry that guarantees that the model output is a real-valued function. To this end, we first observe that an exchange of the multi-indices $\boldsymbol{j}$ and $\boldsymbol{k}$ yields the complex conjugate of the original term:

$$[\gamma_j^* \gamma_k M_{j,k} e^{i\boldsymbol{x} \cdot (\lambda_k - \lambda_j)}]^* = \gamma_k^* \gamma_j M_{j,k}^* e^{i\boldsymbol{x} \cdot (\lambda_j - \lambda_k)} \tag{C10}$$

$$= \gamma_k^* \gamma_j M_{k,j} e^{-i\boldsymbol{x} \cdot (\lambda_k - \lambda_j)}. \tag{C11}$$

Other than that, the coefficients can be freely chosen. To this end, we fix our initial state as the equal superposition state which can be prepared by applying a Hadamard gate to every qubit in the system. This gives $\gamma_j = 1/\sqrt{2^{Nd}}$ and results in the model

$$f_{m'}(\boldsymbol{x}) = 2^{-Nd} \sum_j \sum_k M_{j,k} e^{i\boldsymbol{x} \cdot (\lambda_k - \lambda_j)}. \tag{C12}$$

With this choice, we see that the coefficients are directly proportional to the different entries of the observable M. Recall that our initial goal was to construct the Fourier series $\tilde{g}$ with coefficients $\{c_{\boldsymbol{n}}\}$ where $\boldsymbol{n} \in \mathbb{Z}_K^N$. We already argued that all those are accessible in the frequency spectrum of our model because of the universal nature of the Hamiltonian family $\{H_m\}$. As any frequency corresponds to one or more pairings of multi-indices $\boldsymbol{j}$ and $\boldsymbol{k}$, we can always select a set of these multi-indices such that it is in one-to-one correspondence with the frequencies present in the Fourier series $\tilde{g}$:

$$I = \{(\boldsymbol{j}, \boldsymbol{k}) \in [2^d]^N \times [2^d]^N | \text{for all } \boldsymbol{n} \in \mathbb{Z}_K^N$$
$$\times \text{ there is exactly one pair } (\boldsymbol{j}, \boldsymbol{k}) \text{ so that } \lambda_j - \lambda_k = \boldsymbol{n}\}. \tag{C13}$$

With this it is now straightforward to use the freedom to choose our observable to fix $f_{m'} = \tilde{g}$ by choosing the diagonal and upper-triangular elements of M via

$$M_{j,k} = \begin{cases} 2^{Nd} c_{\boldsymbol{n}} \text{ if } \lambda_j - \lambda_k = \boldsymbol{n} \text{ and } (\boldsymbol{j}, \boldsymbol{k}) \in I, \\ 0 \text{ otherwise}, \end{cases} \tag{C14}$$

after which the lower-triangular elements are fixed by the constraint that the observable is Hermitian. $\blacksquare$

[1] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

[2] J. Romero and A. Aspuru-Guzik, Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions, Adv. Quantum Technol. **4**, 2000003 (2021).

[3] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).

[4] E. Farhi and H. Neven, Classification with quantum neural networks on near term processors, arXiv:1802.06002.

[5] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[6] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Sci. Technol. **4**, 043001 (2019).

[7] M. Ostaszewski, E. Grant, and M. Benedetti, Structure optimization for parameterized quantum circuits, Quantum **5**, 391 (2021).

[8] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).

[9] G. Verdon, J. Marks, S. Nanda, S. Leichenauer, and J. Hidary, Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm, arXiv:1910.02071.

[10] G. Verdon, T. McCourt, E. Luzhnica, V. Singh, S. Leichenauer, and J. Hidary, Quantum graph neural networks, arXiv:1909.12264.

[11] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nat. Phys. **15**, 1273 (2019).

[12] J.-G. Liu and L. Wang, Differentiable learning of quantum circuit born machines, Phys. Rev. A **98**, 062324 (2018).

[13] A. Harrow and J. Napp, Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms, arXiv:1901.05374.

[14] C. Ciliberto, A. Rocchetto, A. Rudi, and L. Wossnig, Statistical limits of supervised quantum learning, Phys. Rev. A **102**, 042414 (2020).

[15] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat Commun **12**, 1791 (2021).

[16] L. Carleson, On convergence and growth of partial sums of Fourier series, Acta Math. **116**, 135 (1966).

[17] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum **4**, 226 (2020).

[18] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Continuous-variable quantum neural networks, Phys. Rev. Res. **1**, 033063 (2019).

[19] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, Adv. Quantum Technol. **2**, 1900070 (2019).

[20] H. Chen, L. Wossnig, S. Severini, H. Neven, and M. Mohseni, Universal discriminative quantum neural networks, Quantum Machine Intelligence **3**, 1 (2021).

[21] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, Expressive power of parametrized quantum circuits, Phys. Rev. Research **2**, 033125 (2020).

[22] J. Biamonte, Universal variational quantum computation, Phys. Rev. A **103**, L030401 (2021).

[23] M. C. Caro and I. Datta, Pseudo-dimension of quantum circuits, Quantum Mach. Intell. **2**, 14 (2020).

[24] M. Schuld and N. Killoran, Quantum Machine Learning in Feature Hilbert Spaces, Phys. Rev. Lett. **122**, 040504 (2019).

[25] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature (London) **567**, 209 (2019).

[26] R. LaRose and B. Coyle, Robust data encodings for quantum classifiers, Phys. Rev. A **102**, 032420 (2020).

[27] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, Quantum embeddings for machine learning, arXiv:2001.03622.

[28] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum Support Vector Machine for Big Data Classification, Phys. Rev. Lett. **113**, 130503 (2014).

[29] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Circuit-centric quantum classifiers, Phys. Rev. A **101**, 032308 (2020).

[30] T. Hubregtsen, J. Pichlmeier, P. Stecher, and K. Bertels, Evaluation of parameterized quantum circuits: On the relation between classification accuracy, expressibility, and entangling capability, Quantum Mach. Intell. **3**, 9 (2021).

[31] N. Koide-Majima and K. Majima, Quantum circuit-like learning: A fast and scalable classical machine-learning algorithm with similar performance to quantum circuit learning, arXiv:2003.10667.

[32] M. Watabe, K. Shiba, M. Sogabe, K. Sakamoto, and T. Sogabe, Quantum circuit parameters learning with gradient descent using backpropagation, arXiv:1910.14266.

[33] C. Blank, D. K. Park, J.-K. K. Rhee, and F. Petruccione, Quantum classifier with tailored quantum kernel, arXiv:1909.02611.

[34] J. Zhao, Y.-H. Zhang, C.-P. Shao, Y.-C. Wu, G.-C. Guo, and G.-P. Guo, Building quantum neural networks based on a SWAP test, Phys. Rev. A **100**, 012334 (2019).

[35] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, K. McKiernan, J. J. Meyer, Z. Niu, A. Száva, and N. Killoran, PennyLane: Automatic differentiation of hybrid quantum-classical computations, arXiv:1811.04968.

[36] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, U.K., 2000).

[37] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms* (Cambridge University Press, Cambridge, U.K., 2014).

[38] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, Understanding deep learning requires rethinking generalization, arXiv:1611.03530.

[39] Y. Jiang, B. Neyshabur, H. Mobahi, D. Krishnan, and S. Bengio, Fantastic generalization measures and where to find them, arXiv:1912.02178.

[40] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, Implicit neural representations with periodic activation functions, arXiv:2006.09661.

[41] A. Zhumekenov, M. Uteuliyeva, O. Kabdolov, R. Takhanov, Z. Assylbekov, and A. J. Castro, Fourier neural networks: A comparative study, arXiv:1902.03011.

[42] S. Wahls, V. Koivunen, H. V. Poor, and M. Verhaegen, Learning multidimensional fourier series with tensor trains, in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (IEEE, New York, 2014), pp. 394–398.

[43] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.103.032430 for a reporting of the carbon costs associated with this work.

[44] F. J. Gil Vidal and D. O. Theis, Input redundancy for parameterized quantum circuits, Front. Phys. **8**, 297 (2020).

[45] https://github.com/XanaduAI/expressive_power_of_quantum_models.

[46] F. Weisz, Summability of multi-dimensional trigonometric Fourier series, Surv. Approximation Theory **7**, 1 (2012), https://www.emis.de/journals/SAT/.